LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-221227

Multi-Model Combination Techniques for Hydrological Forecasting: Application to Distributed Model Intercomparison Project Results

N. Ajami, Q. Duan, X. Gao, S. Sorooshian

May 9, 2006

Journal of Hydrometeorology

Multi-Model Combination techniques for Hydrological Forecasting: Application to Distributed Model Intercomparison Project Results

Newsha k. Ajami[1,*], Qingyun Duan[2], Xiaogang Gao[1], Soroosh Sorooshian[1]

1. University of California at Irvine (UCI), Irvine, CA
2. Lawrence Livermore National Laboratory, Livermore, CA

Abstract

This paper examines several multi-model combination techniques: the Simple Multi-model Average (SMA), the Multi-Model Super Ensemble (MMSE), Modified Multi-Model Super Ensemble (M3SE) and the Weighted Average Method (WAM). These model combination techniques were evaluated using the results from the Distributed Model Intercomparison Project (DMIP), an international project sponsored by the National Weather Service (NWS) Office of Hydrologic Development (OHD). All of the multi-model combination results were obtained using uncalibrated DMIP model outputs and were compared against the best uncalibrated as well as the best calibrated individual model results. The purpose of this study is to understand how different combination techniques affect the skill levels of the multi-model predictions. This study revealed that the multi-model predictions obtained from uncalibrated single model predictions are generally better than any single member model predictions, even the best calibrated single model predictions. Furthermore, more sophisticated multi-model combination techniques that incorporated bias correction steps work better than simple multi-model average predictions or multi-model predictions without bias correction.

* Corresponding author address: Newsha K. Ajami, University of California at Irvine, Dept. of Civil and Environmental Eng., E/4130 Engineering Gateway, Irvine, CA, 82697-2175.
E-mail address: nkhodata@uci.edu

1. Introduction:

Many hydrologists have been working to develop new hydrologic models or to try improving the existing ones. Consequently, a plethora of hydrologic models are in existence today, with many more likely to emerge in the future (Singh 1995, Singh and Frevert, 2002a and 2002b). With the advancement of the Geographic Information System (GIS), a class of models, known as distributed hydrologic models, has become popular (Russo et al., 1994, Vieux, 2001). These models explicitly account for spatial variations in topography, meteorological inputs and water movement. The National Weather Service Hydrology Laboratory has recently conducted the Distributed Model Intercomparison Project (DMIP) that showcased the state-of-the-art distributed hydrologic models from different modeling groups (Smith et al., 2004). It was found that there is a large disparity in the performance of the DMIP models (Reed et al., 2004). The more interesting findings were that multi-model ensemble averages perform better than any single model predictions, including the best calibrated single model prediction, and that multi-model ensemble averages are more skillful and reliable than the single model ensemble averages (Georgakakos et al., 2004). Georgakakos et al. (2004) attributed the superior skill of the multi-model ensembles to the fact that model structural uncertainty is accounted for in the multi-model approach. They went on to suggest that multi-model ensemble predictions should be considered as an operational forecasting tool. The fact that the simple multi-model averaging approach such as the one used by Georgakakos et al. (2004) has led to more skillful and reliable predictions has motivated us to examine whether more sophisticated multi-model combination techniques can result in consensus predictions of even better skills.

Most hydrologists are used to the traditional contructionist approach, in which the goal of the modeler is to build a perfect model that can capture the real world processes as much as possible. Multi-model combination approach, on the other hand, works in essentially a different paradigm in which the modeler aims to extract as much information as possible from the existing models. The idea of combining predictions from multiple models was explored more than thirty years ago in econometrics and statistics (see Bates and Granger, 1969; Dickinson, 1973 and 1975; Newbold and Granger, 1974). In 1976, Thompson applied the model combination concept in weather forecasting. He showed that the mean square error of forecast generated by combining two independent model outputs is less than that of the individual predictions. Based on the study done by Clement (1989), the concept of the combination forecasts from different models were applied in diverse fields ranging from management to weather prediction. Fraedrich and Smith (1989) presented a linear regression technique to combine two statistical forecast methods for long-range forecasting of the monthly tropical Pacific sea surface temperatures (SST). Krishnamurti et al. (1999) explored the model combination technique by using number of forecasts from a selection of different weather and climate models. They called their technique Multi-Model Superensemble (MMSE) and compared it to simple model averaging (SMA) method. Krishnamurti and his group applied the MMSE technique to forecast various weather and climatological variables (e.g. precipitation, tropical cyclones, seasonal climate) and all of these studies agreed that consensus forecast outperforms any single member model as well as the SMA technique (e.g. Krishnamurti et al., 1999; Krishnamurti, et al., 2000a,b; Krishnamurti et al., 2001; Krishnamurti et al., 2002; Mayers et al., 2001; Yun et al. 2003). Khrin and

Zwiers (2002) reported that for small sample size data the MMSE does not perform as well as simple ensemble mean or the regression-improved ensemble mean.

Shamseldin et al, (1997) first applied the model combination technique in the context of rainfall-runoff modeling. They studied three methods of combining model outputs, the SMA method, the Weighted Average Method (WAM) and the Artificial Neural Network (ANN) method. They applied these methods to combine outputs of five rainfall-runoff models for eleven watersheds. For all these cases they reported that the model combination prediction is superior to that of any single model predictions. Later Shamseldin and O'Connor (1999) developed a Real-Time Model Output Combination Method (RTMOCM), based on the Linear Transfer Function Model (LTFM) and the WAM and tested it using three rainfall-runoff models on five watersheds. Their results indicated that the combined flow forecasts produced by RTMOCM were superior to those from the individual rainfall-runoff models. Xiong et al. (2001) refined the RTMOCM method by introducing the concept of Takagi-Sugeno fuzzy system as a new combination technique. Abrahart and See (2002) compared six different model combination techniques: the SMA; a probabilistic method in which the best model from the last time step is used to create the current forecast; two different neural network operations and two different soft computing methodologies. They found that neural network combination techniques perform the best for a stable hydro-climate regime, while fuzzy probabilistic mechanism generated superior outputs for more volatile environment (flashier catchments with extreme events).

This paper extends the work of Georgakakos et al. (2004) and that of Shamseldin et al. (1997) by examining several multi-model combination techniques, including SMA, MMSE, WAM, and a variant of MMSE, known as Bias Corrected Multi-model Average (BCMA). As in Georgakakos et al. (2004), we will use the results from DMIP to evaluate various multi-model combination techniques. Through this study, we would like to answer the following basic question: "Does it matter which multi-model combination techniques are used to obtain consensus prediction"? We will also investigate how the skills of the multi-model predictions are influenced by different factors, including the seasonal variations of hydrological processes, number of independent models considered, lengths of training data, etc. The paper is organized as follows. Section 2 overviews different model combination techniques. Section 3 describes the data used in this study. Section 4 presents the results and analysis. Section 5 provides major lessons and conclusions.

2. Brief Description of the Multi-model Combination Techniques

2.1 *Multi-Model SuperEnsemble, MMSE:*

Multi-Model Super-Ensemble, MMSE, is a multi-model forecasting approach popular in meteorological forecasting. MMSE uses the following logic (Krishnamurti et al., 2000):

$$(Q_{MMSE})_t = \overline{Q}_{obs} + \sum_{i=1}^{N} x_i((Q_{sim})_{i,t} - (\overline{Q}_{sim})_i) \quad (1)$$

Where $(Q_{MMSE})_t$ is the multi-model prediction obtained through MMSE at time *t*, $(Q_{sim})_{i,t}$ is the *i*th model streamflow simulation for time t, $(\overline{Q}_{sim})_i$ is the average of the *i*th model prediction over the training period, $(\overline{Q}_{obs})$ is observed average over the training period, $\{x_i, i=1,2,\ldots, N\}$are the regression coefficients (weights) computed over the training period, and finally *N* is the number of hydrologic models.

Equation (1) comprises two main terms. First term, $(\overline{Q}_{obs})$, which replaces the MMSE prediction average with the observed average, serves to reduce the forecast bias. Second term $\sum x_i[(Q_{sim})_{i,t} - (\overline{Q}_{sim})_i]$, reduces the variance of the combination predictions, using multiple regressions. Therefore, the logic behind this methodology is a simple idea of bias correction along with variance reduction. We should also note that when a multi-model combination technique such as MMSE is used to predict hydrologic variables like river flows, it is important that the average river flows during the training period over which the model weights are computed should be close to the average river flow of the prediction period (i.e., the stationarity assumption). In Section 4, we will show that bias removal and stationarity assumption are important factors in multi-model predictive skills.

2.2. *Modified Multi-Model Super Ensemble, M3SE*

Modified Multi–Model Super Ensemble (M3SE) technique is a variant of the MMSE. This technique works in the same way as in MMSE except the bias correction step. In MMSE, model bias is removed by replacing the average of the predictions by the

average of observed flows. In M3SE, the bias is removed by mapping the model prediction at each time step to the observed flow with the same frequency as the forecasted flow. Figure (1) illustrates how forecasted flows are mapped into observed flows through frequency mapping. The solid arrow shows the original value of the forecast and the dashed arrow points to the corresponding observed value. The frequency mapping bias correction method has been popular in hydrology because it the bias corrected hydrologic variables agree well statistically with the observations, while the bias correction procedure used in MMSE might lead to unrealistic values (i.e., negative values). After removing bias from each model forecast, the same solution procedure for MMSE is applied to M3SE.

2.3. *Weighted Average method, WAM*

Weighted Average Method (WAM) is one of the model combination techniques specifically developed for rainfall-runoff modeling by Shamseldin et al. (1997). This method also utilizes the Multiple Linear Regression (MLR) technique to combine the model predictions. The model weights are constrained to be always positive and to sum up to unity. If we have model predictions from N models, WAM can be expressed as:

$$(Q_{WAM})_t = \sum_{i=1}^{N} x_i \cdot (Q_{sim})_{i,t} \qquad (2)$$

$$S.t.$$

$$\begin{cases} x_i > 0 \\ \sum x_i = 1 \quad i = 1....N \end{cases}$$

Where $(Q_{WAM})_t$ is the multi-model prediction obtained through WAM at time *t*. Constrained Least Square can be used to solve the equation and estimate the weights. For more details about this method reader should refer to Shamseldin et al. (1997).

2.4 Simple Model Average, SMA

The Simple Model Average (SMA) method is the multi-model ensemble technique used by Georgakakos et al. (2004). This is the simplest technique and is used as a benchmark in evaluating more sophisticated techniques in this work. SMA can be expressed by the following equation:

$$(Q_{SMA})_t = \overline{Q}_{obs} + \sum_{i=1}^{N} \frac{(Q_{sim})_{i,t} - (\overline{Q}_{sim})_i}{N} \tag{3}$$

Where $(Q_{SMA})_t$ is the multi-model prediction obtained through SMA at time *t*.

2.5 Differences Between the Four Multi-model Combination Techniques

The major differences between these multi-model combination methods are the model weighting scheme and the bias removal scheme. MMSE, M3SE and WAM have variable model weights, while SMA has equal model weights. MMSE and M3SE compute the model weights through multiple linear regressions while WAM computes the model weights using constrained least square approach that ensures positive model weights and total weights equal to 1. With respect to bias correction, MMSE and SMA remove the bias by replacing the prediction mean with the observed mean, while WAM does not incorporate any bias correction. M3SE removes the bias by using frequency mapping method as illustrated in Section 2.3.

3. The Study Basins and Data:

We have chosen to evaluate the multi-model combination methods using model outputs collected from DMIP (Smith et al., 2004). DMIP was conducted over the basins in the Arkansas Red River basins. Five basins of the DMIP basins are included in this study: Illinois River basin at Watts, OK, Illinois River basin at Eldon, OK, Illinois River basin at Tahlequah, OK, Blue River basin at Blue, OK, and Elk River basin at Tiff City, MO. Fig. 2 shows the location of the basins while Table 1 lists the basin topographic and climate information. Silty clay is the dominant soil texture type of those basins, except for Blue River, where the dominant soil texture is clay. The land cover of those basins is dominated by natural forest and agriculture crops (Smith et al., 2004).

The average maximum and minimum surface air temperature in the region are approximately 22°C and 9°C, respectively. Summer maximum temperatures can get as high as 38°, and freezing temperatures occur generally in December through February. The climatological annual average precipitation of the region is between 1010-1160 mm/yr (Smith et al., 2004).

Seven different modeling groups contributed to DMIP by producing flow simulation for the DMIP basins using their own distributed models, driven by DMIP provided meteorological forcing data. The precipitation data, available at 4x4 km^2 spatial resolution, was generated from the NWS Next-generation Radar (NEXRAD). Other meteorological forcing data such as air temperature, downward solar radiation, humidity and wind speed were obtained from the University of Washington (Maurer et al., 2001).

Table 2 lists the participating groups and models. For more details on model description and simulation results, readers should refer to Reed et al. (2004).

For this study, we obtained the river flow simulations from all participating models for the entire DMIP study period: 1993-1999. The uncalibrated river simulation results are used for multi-model combination study. Observed river flow data, along with the best calibrated single model flow simulations from the DMIP, are used as the benchmarks for comparing skill levels of the different multi-model predictions. Unless otherwise specified, data period from 1993 to 1996 was used to train the model weights from the multi-model combination techniques, while the rest of the data period (1997-1999) was used for validating the consistency of the multi-model predictions using these weights.

4. Multi-model Combination Results and Analysis

4.1 Model evaluation criteria

Before we present the results, two different statistical criteria are introduced: the Hourly Root Mean Square Error (HRMS) and the Pearson correlation coefficient (R). These criteria are used to compare the skill levels of different model predictions. These criteria are defined as follows:

$$HRMS = \sqrt{(\frac{1}{n}\sum_{t=1}^{n}((Q_{sim})_t - (Q_{obs})_t)^2} \quad (3)$$

$$R = \frac{\sum_{t=1}^{n}((Q_{obs})_t (Q_{sim})_t) - [n\overline{Q}_{obs}\overline{Q}_{sim}]}{\sqrt{[\sum_{t=1}^{n}(Q_{obs})_t^{\,2} - n(\overline{Q}_{obs})^2][\sum_{t=1}^{n}(Q_{sim})_t^{\,2} - n(\overline{Q}_{sim})^2]}} \quad (4)$$

4.2. Comparison of the Multi-model Consensus Predictions and the Uncalibrated Individual Model Predictions

In the first set of numerical experiments, the multi-model predictions were computed from the uncalibrated individual model predictions using different multi-model combination techniques described in Section 2. Figures 3a and 3b compare the HRMS and R values of the individual model predictions against those of the SMA predictions. The horizontal axis in Figures 3a and 3b denotes the statistics from the individual models, while the vertical axis denotes that from the SMA predictions. These figures clearly show that the statistics from the individual model predictions are worse than those of the SMA predictions. These results are totally consistent with the conclusions from the paper by Georgakakos et al. (2004).

Figures 4 and 5 show the comparison results of the different multi-model combination techniques against each other and against the best uncalibrated individual model predictions during the training and validation periods. The horizontal axis denotes the different multi-model predictions, along with the best individual model predictions. Clearly shown in these figures is that all multi-model predictions have superior performance statistics compared to the best individual model predictions. More interestingly, the multi-model predictions generated by MMSE and M3SE show noticeably better performance statistics than those by SMA. This implies that there are

indeed benefits in investigating more sophisticated multi-model combination techniques. The predictions generated by WAM show worse performance statistics than the predictions generated by other multi-model combination techniques. This suggests that the bias removal step incorporated by other multi-model combination techniques is important in improving predictive skills.

The obvious advantage of multi-model predictions from the training period carries into the validation period in almost all cases except for Blue River basin, where the performance statistics of the multi-model predictions are equal to or slightly worse than the best individual model predictions. The reason for the relative poor performance in Blue River basin is that a noticeable change in flow characteristics is observed from the training period to the validation period (i.e., the average flow changes from 10.8cms in the training period to 7.17cms in the validation period, standard deviation from 27.6cms to 16.8cms). This indicates that the stationarity assumption for river flow was violated. Consequently the skill levels of the predictions during validation period were adversely affected.

According to Reed et al. (2004), the calibrated model predictions from the distributed model operated by NWS OHD (hereafter, denoted as OHD-cal) have the best performance statistics. To get a measure of how multi-model predictions fare against the best calibrated single model predictions, Figures 6a and 6b show the scatter plots of the HRMS and R for all multi-model combination techniques as well as for OHD-cal for the training and validation periods. As revealed in the figures, MMSE and M3SE outperform the OHD-cal for all the basins except Blue River Basin during the training period.

During validation period, however, OHD-cal has shown a slight advantage in performance statistics over the multi-model predictions. MMSE and M3SE are shown to be the best performing combination technique during validation period and have statistics closer to those of the OHD-cal, while WAM and SMA have worse performance statistics.

4.3. Application of Multi-model Combination Techniques to River Flow Predictions from Individual Months

Hydrological variables such as river flows are known to have a distinct annual cycle. The predictive skills of hydrologic models for different months often mimic this annual cycle, as shown in Figure 7 which displays the performance statistics of the individual model predictions for Illinois River basin at Eldon (during the training period?). Figure 7 reveals that a model might perform well in some months, but poorly in other months, when compared to other models. This led us to hypothesize that the weights for different months should take on different sets of values to obtain consistently skillful predictions for all months. To test this hypothesis, we applied multi-model combination techniques to flow values from each individual month separately. Model weights for each calendar month were computed separately for all basins and all multi-model combination techniques.

Figures 8 and 9 show the comparison of HRMS and R statistics of all combination techniques applied to entire training periods and to individual months during training and validation periods. Also shown is the statistics for OHD-cal. From the figures, it is clear that the performance of combination techniques with monthly weights is generally better than that of combination techniques with single sets of weights for the entire training

period. During the validation period, however, the performance statistics using single sets of weights are generally better than those using monthly weights. This is because that the stationarity assumptions are more easily violated when the multi-model techniques are applied monthly.

4.4. The Effect of Different Number of Models Used for Model Combination on Predictive Skills

One often asked question on multi-model predictions is how many models are needed to ensure good skills from multi-model predictions. To address this question, we performed a series of experiments by sequentially removing different number of models from consideration. Figure 10 **(create this figure)** displays the test results. Shown in the figure are the average HRMS and R statistics when different number of models were included in model combination. **(add more discussion based on the actual figure)** To illustrate how important the skills of individual models are on the skills of the multi-model predictions, we experimented with removing the best performing model and the worst performing model from consideration. The results are also shown in Figure 10. It is clear that excluding the best model would deteriorate the predictive skills more significantly compared to eliminating the weakest model.

5. Conclusion and future direction

We have applied four different multi-model combination techniques to the multi-model results from the DMIP, an international project sponsored by NWS Office of Hydrologic Development to intercompare seven state-of-the-art distributed hydrologic

models in use today (Smith et al., 2004). This work is motivated by the fact that despite the progress in hydrologic model development, models still do not agree with each other. Developing more sophisticated models may lead to more agreement among models. Taking advantage the strengths of the existing models may be more profitable.

We have learned several valuable lessons from this work. First, simply averaging the individual model predictions would result in consensus multi-model predictions that are superior to any single member model predictions. More sophisticated multi-model combination approaches such as MMSE and M3SE can improve the predictive skills even further. The results obtained here show that the multi-model predictions generated by MMSE and M3SE are even better than or at least are comparable to the best calibrated single model predictions. This suggests that future operational hydrologic predictions should incorporate multi-model prediction strategy.

Second, in examining the different multi-model combination strategy, it was found that bias removal is an important step in improving the predictive skills of the multi-model predictions. MMSE and M3SE predictions, which incorporated bias correction steps, perform noticeably better than WAM predictions, which did not. Also important is the stationarity assumption when using multi-model combination techniques for predicting hydrologic variables such as river flows. In Blue River basin where the average river flow values are significantly different between the training and validation periods, the advantages of multi-model predictions was lost during the validation period. This finding was also confirmed when the multi-model combination techniques were applied to river flows from individual months.

Third, we attempted to address how many models are needed to ensure the good skills of multi-model predictions. We found that at least (Insert more discussion based on the figure). We also found that the multi-model prediction skills are related to the skills of the individual member models. If the prediction skill from an individual model is poor, removing this model from consideration does not affect the skill of the multi-model predictions very much. On the other hand, removing the best model from consideration does adversely affect the multi-model prediction skill.

This work was based on a limited data set. There are only seven models and a total of seven years of data. The findings are necessarily subject to these limitations. The regression based techniques used here (i.e., MMSE, M3SE and WAM) are vulnerable to multi-colinearity problem that may result in unstable or unreasonable estimates of the weights (Winkler, 1989). This in turn would reduce the substantial advantages achieved employing these combination strategies. There are remedies available to deal with colinearity problem (Shamseldin, et al., 1997; Yun et al., 2003). This may entail more independent models to be included in the model combination.

Multi-linear regression based approach presented here is only one type of the multi-model combination approach. Over recent years, there are other model combination approaches developed in fields other than hydrology, such as the Bayesian Model Average (BMA) method, in which model weights are proportional to the individual model skills and can be computed recursively as more observation information become available (Hoeting et al., 1998?). Model combination techniques are still young

in hydrology. The results presented in this paper and other papers show promise that multi-model predictions will be a superior alternative to current single model prediction.

Acknowledgment

This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

References

Abrahrt, R.J., See, L., 2002. Multi –model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments. Hydrology and Earth System Sciences, 6(4), 655-670.

Ajami, N.K., Gupta, H., Wagener, T., Sorooshian, S., 2004. Calibration Of A Semi-Distributed Hydrologic Model For Streamflow Estimation Along A River System. Journal of Hydrology, 298(1-4), 112-135.

Bates, J.M., and Granger, C.W.J., 1969. The combination of forecasts, Operational Research Quarterly, 20, 451-468.

Beven, K., and Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249, 11-29.

Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. International Journal of Forecasting, 5, 559-583.

Crawford, N.H. and Linsley, R. K., 1966. Digital Simulation in Hydrology – Stanford Watershed Model IV. Technical Report No. 39, Department of Civil and Environmental Engineering, Stanford University, Stanford, California.

Dickinson, J.P., 1973. Some statistical results in the combination of forecast. Operational Research Quarterly, 24(2), 253-260.

Dickinson, J.P., 1975. Some comments on the combination of forecasts. Operational Research Quarterly, 26, 205-210.

DMIP website, 2001. http://www.nws.noaa.gov/oh/hrl/dmip. Accessed May. 2004.

Fraedrich, K., and Smith, N.R., 1989. Combining predictive schemes in long-range forecasting . Journal of Climate, 2, 291-294.

Georgakakos, K.P., Seo, D.J., Gupta, H. Schake J. and Butts, M. B., 2004. Characterizing streamflow simulation uncertainty through multimodel ensembles. Journal of Hydrology, 298(1-4), 222-241.

Hoffman, R.N., and Kalnay, E., 1983. Lagged average forecasting, an alternative to Monte-Carlo forecasting. Tellus Series a-Dynamic Meteorology and Oceanography 35 (2), 100-118.

Hogue, T.S., Sorooshian, S., Gupta, V.K., Holz, A., and Braatz, D., 2000. A multistep automatic calibration scheme for river forecasting models. Journal of Hydrometeorology, 1, 524-542.

Kharin. V.V., Zwiers, F.W., 2002. Climate predictions with multimodel ensembles. Journal of Climate 15 (7), 793-799.

Krishnamurti, T.N., Kishtawal, C.M., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C.E., Gadgil, S. and Surendran, S., 1999. Improved skill of weather and seasonal climate forecasts from multimodel superensemble. Science 285 (5433), 1548-1550.

Krishnamurti, T.N., Kishtawal, C.M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, 2000. Multimodel ensemble forecasts for weather and seasonal climate. Journal of Climate 13, 4196-4216.

Maurer, E.P., G.M. O'Donnell, D.P. Lettenmaier, and J.O. Roads, 2001, Evaluation of the Land Surface Water Budget in NCEP/NCAR and NCEP/DOE Reanalyses using an Off-line Hydrologic Model. J. Geophys. Res., 106(D16), 17,841-17,862

Mayers, M., Krishnamurti, T.N., Depradine, C., Moseley, L., 2001. Numerical weather prediction over the Eastern Caribbean using Florida State University (FSU) global and regional spectral models and multi-model/multi-analysis super-ensemble. Meteorology and Atmospheric Physics 78 (1-2), 75-88.

Newbold, P., and Granger, C.W.J., 1974. Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society. A., 137(part 2), 131-146.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.J., DMIP participants, 2004. Overall distributed modeling intercomparison project results. Journal of Hydrology, 298(1-4), 27-60.

Russo, R., A. Peano, I. Becchi and G.A. Bemporad, 1994, Advances in Distributed Hydrology, Water Resources Publications, Chelsea, MI, USA, 416p.

Shamseldin, A.Y., Nasr A.E., O'Connor, K.M., 2002. Comparison of the multi-layer feed-forward Neural Network method used for river flow forecasting. Hydrology and Earth System Sciences, 6(4), 671-684.

Shamseldin, A.Y., O'Connor, K.M., Liang, G.C., 1997. Methods for combining the outputs of different rainfall-runoff models. Journal of Hydrology, 197, 203-229.

Shamseldin, A.Y., and O'Connor K.M., 1999. A real-time combination method for the outputs of different rainfall-runoff models. Hydrological Science Journal, 44(6), 895-912.

Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The distributed model intercomparison project (DMIP): an overview. Journal of Hydrology. 298(1–4), 4–26.

Thompson, P.D., 1976. How to improve accuracy by combining independent forecasts. Monthly Weather Review, 105, 228-229.

Toth, Z. and Kalnay, E., 1993. Ensemble forecasting at NMC - the generation of perturbations. Bulletin of the American Meteorological Society 74 (12), 2317-2330.

Vieux, B.E., 2001, Distributed Hydrologic Modeling Using GIS, Kluwer Academic Publishers, Dordrecht, The Netherlands, 294p

Yun, W.T., Stefanova, L. and Krishnamurti, T.N., 2003. Improvement of the multimodel superensemble technique for seasonal forecasts. Journal of Climate 16, 3834-3840.

Table 1. Basin Information

(Create)

Participant	Model	Primary Application	Spatial unit for rainfall-runoff calculation	Rainfall-runoff scheme	Channel routing scheme
Agricultural Research Services (ARS)	SWAT	Land Management/Agricultural	Hydrologic Response Unit (HRU)	Multi-layer soil water balance	Muskingum or Variable storage
University of Arizona (ARZ)	SAC-SMA	Streamflow Forecasting	Sub-basins	SAC-SMA	Kinematic Wave
Environmental Modeling Center (EMC)	NOAH Land Surface Model	Land-atmosphere interactions	1/8 degree grids	Multi-layer Soil water and energy balance	--
Hydrologic Research Center (HRC)	HRCDHM	Streamflow Forecasting	Sub-basins	SAC-SMA	Kinematic Wave
Office of Hydrologic Development (OHD)	HL-RMS	Streamflow Forecasting	16 km^2 grid cells	SAC-SMA	Kinematic Wave
Utah State University (UTS)	TOPNET	Streamflow Forecasting	Sub-basins	TOPMODEL	--
University of Waterloo, Ontario (UWO)	WATFLOOD	Streamflow Forecasting	1-km grid		Linear Storage Routing

Table 2. DMIP participant modeling groups and characteristics of their distributed hydrological models (Reed et al., 2004)

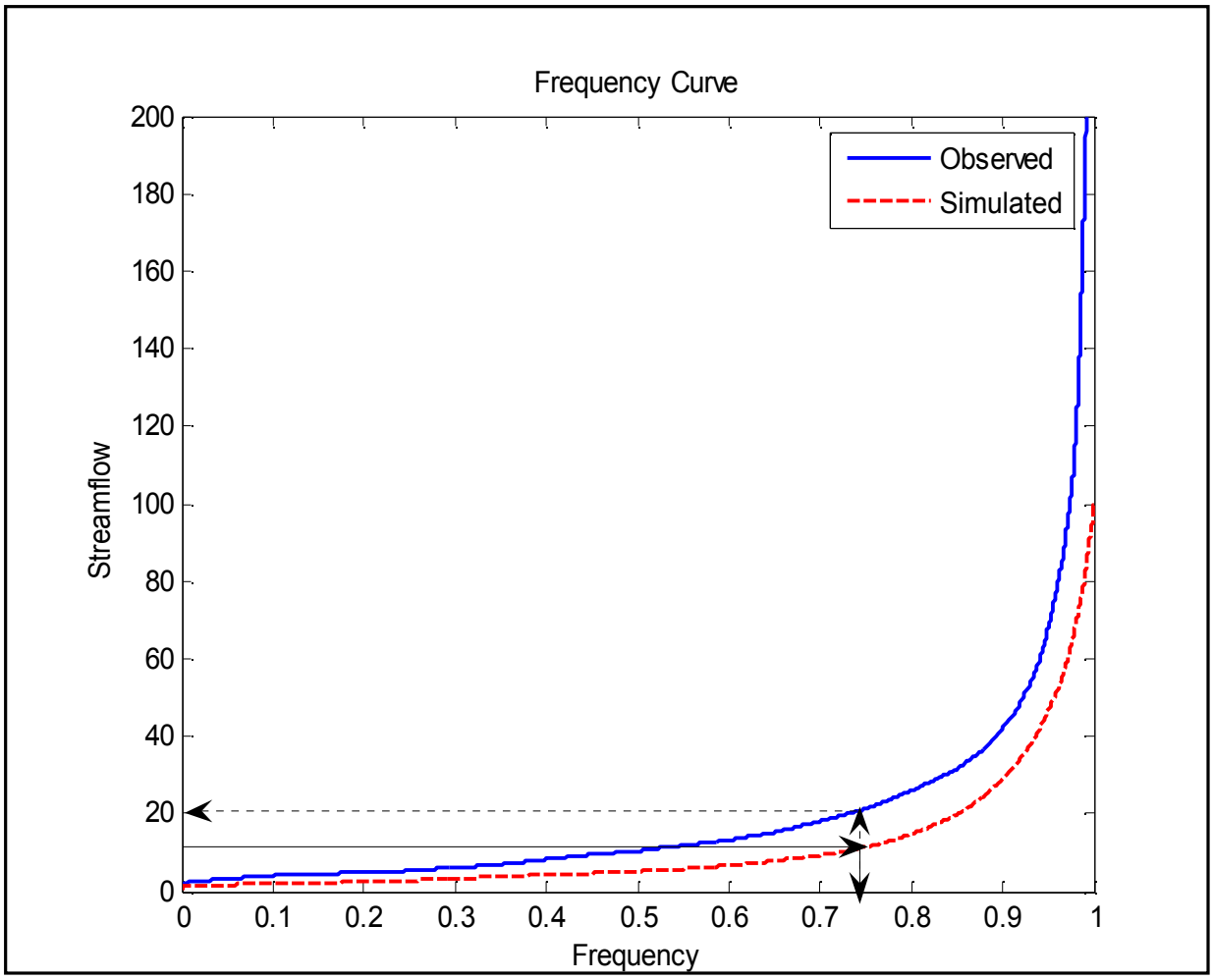


Figure 1: Frequency curve which is being used for Bias-correction for MMC method

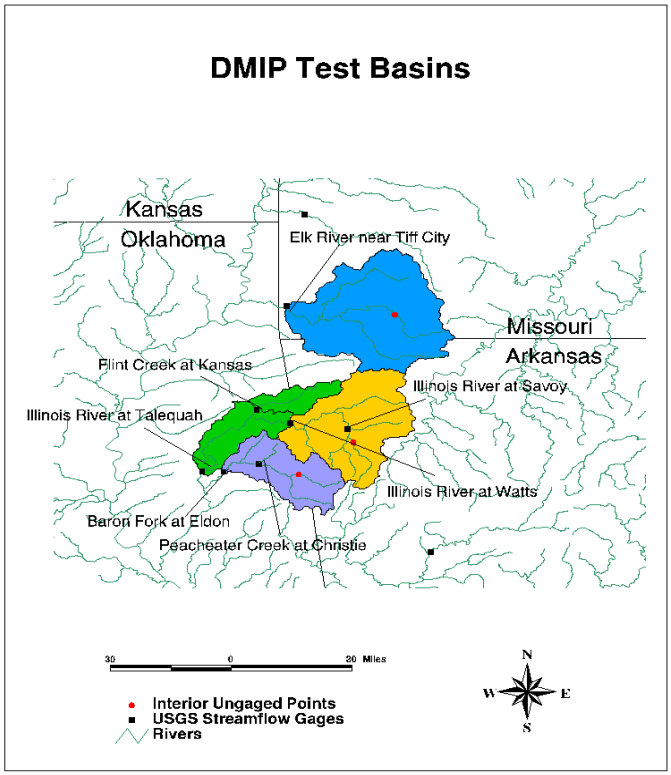
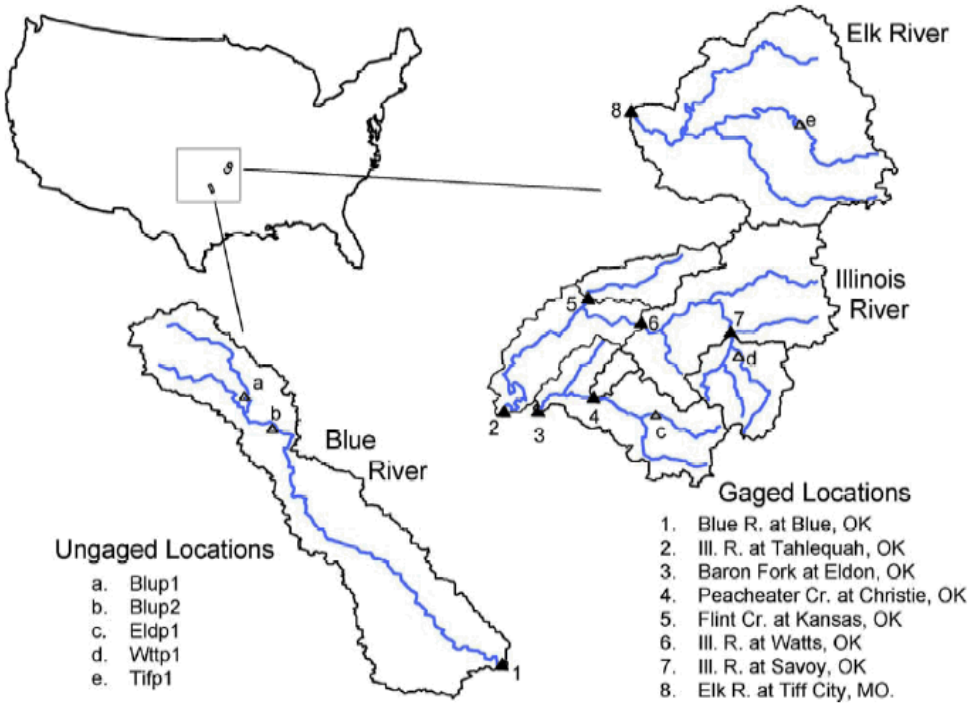


Figure 2. DMIP Test Basins; Circled one is the Illinois River basin with the outlet at Watts. (Source: DMIP website, 2001)


Elk River
Illinois River
Blue River
Ungaged Locations
a. Blup1
b. Blup2
c. Eldp1
d. Wttp1
e. Tifp1
Gaged Locations
1. Blue R. at Blue, OK
2. Ill. R. at Tahlequah, OK
3. Baron Fork at Eldon, OK
4. Peacheater Cr. at Christie, OK
5. Flint Cr. at Kansas, OK
6. Ill. R. at Watts, OK
7. Ill. R. at Savoy, OK
8. Elk R. at Tiff City, MO.

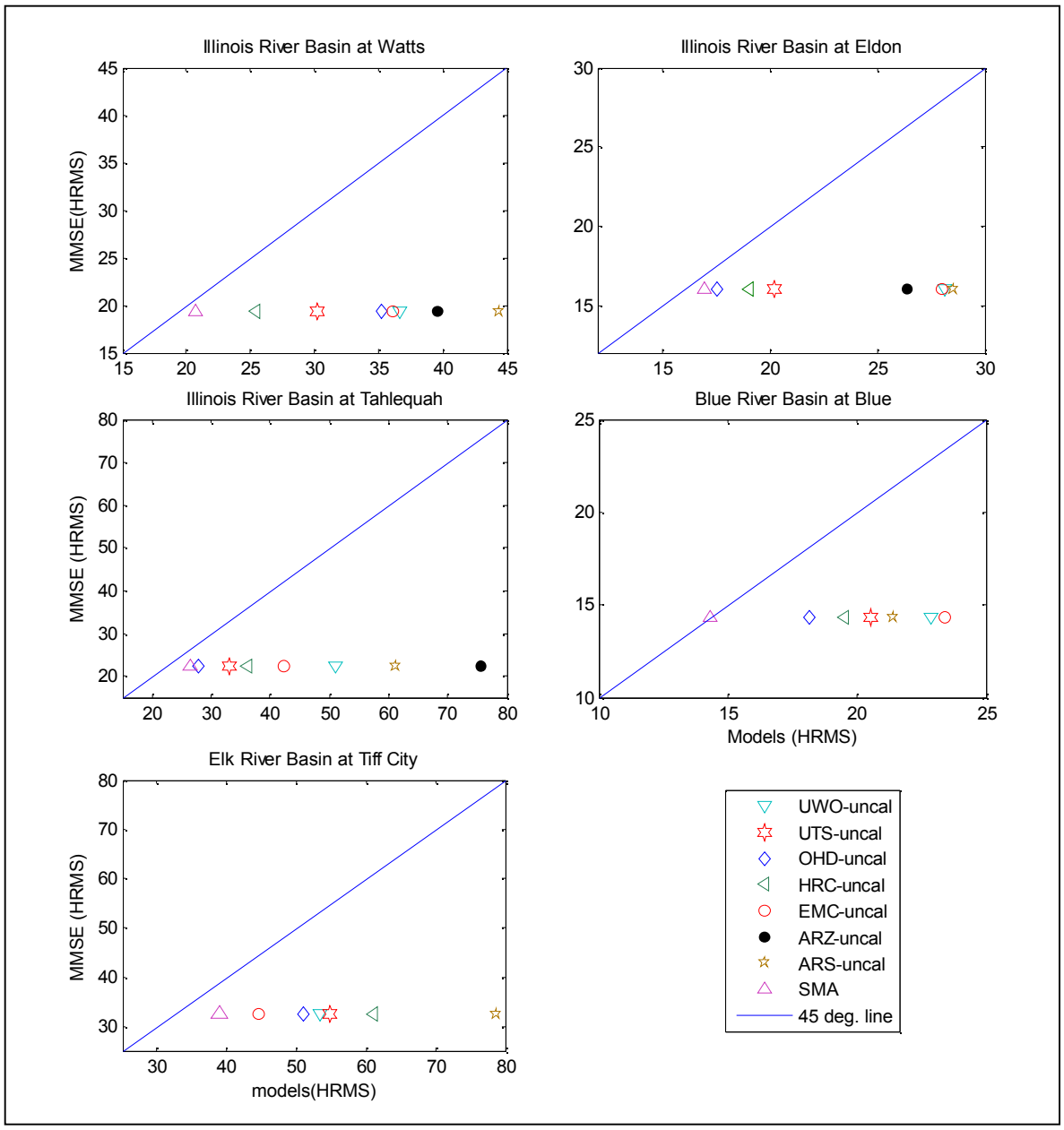


Figure 3. Hourly root mean square error for MMSE versus uncalibrated member models for all the basins.

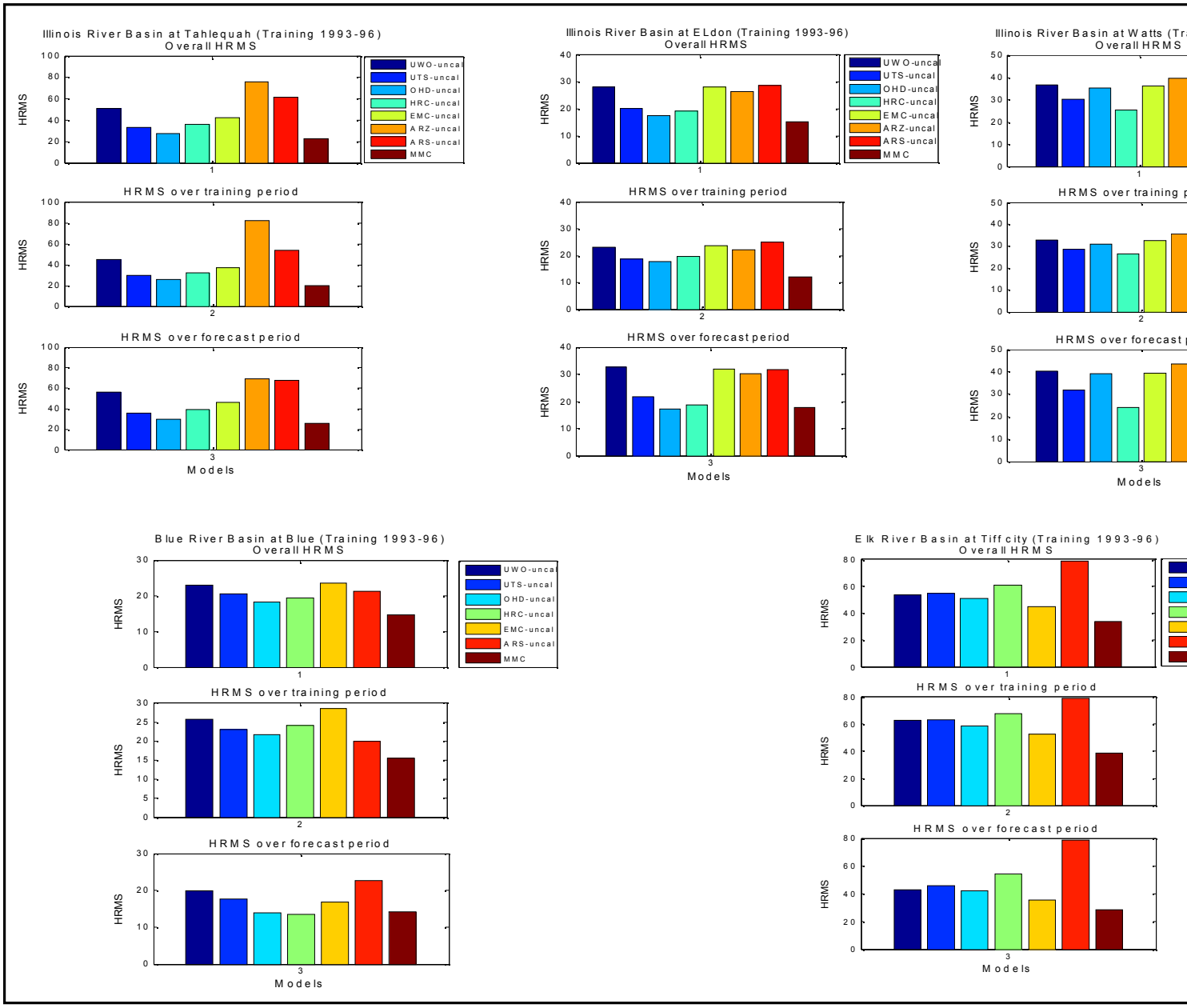


Figure 4.a. Statistical comparison of the MMC's performance (using uncalibrated memebr models) to the skill of any individua
the basins.

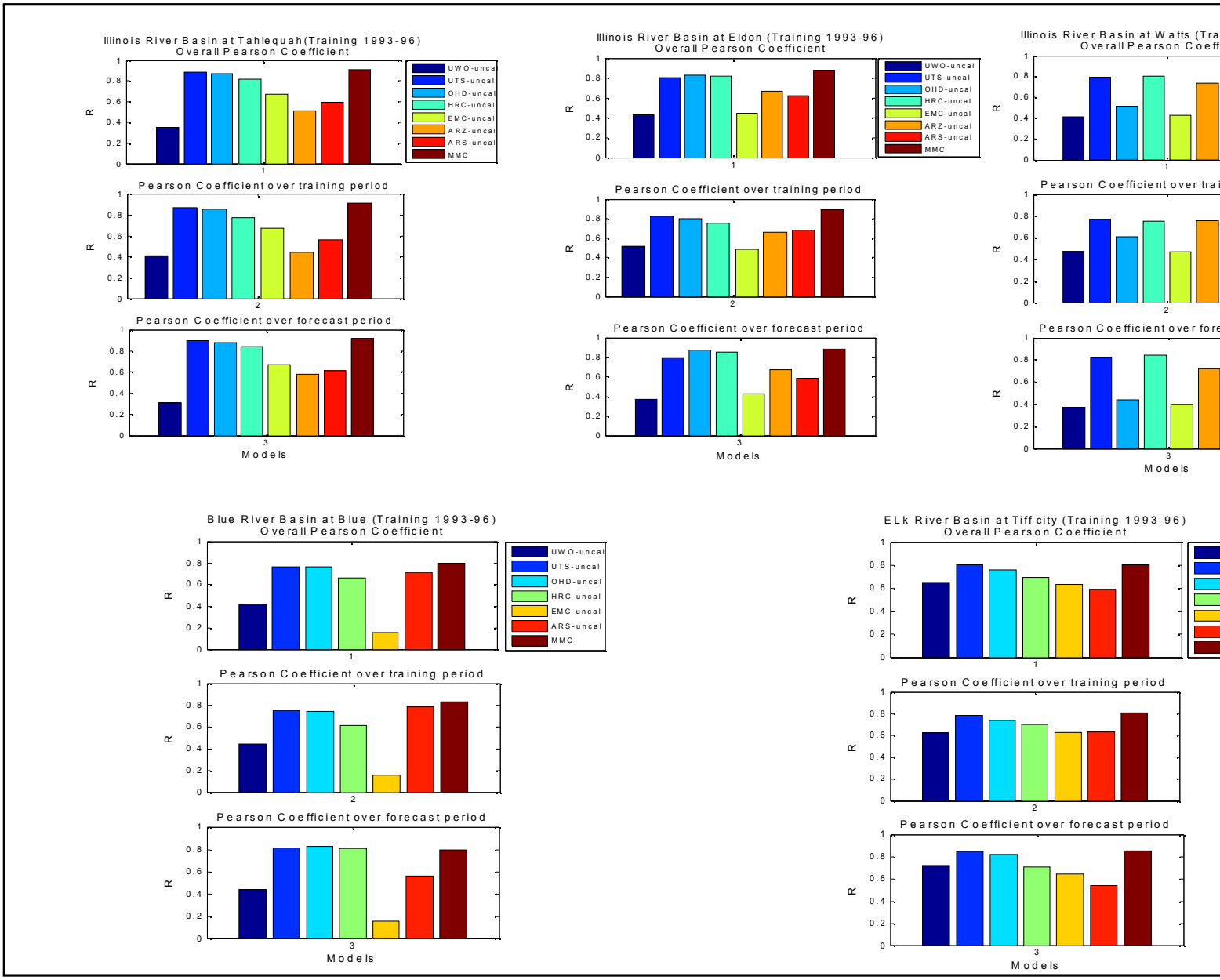


Figure 4.b. Statistical comparison of the MMC's performance (using uncalibrated memebr models) to the skill of any individual m
the basins.

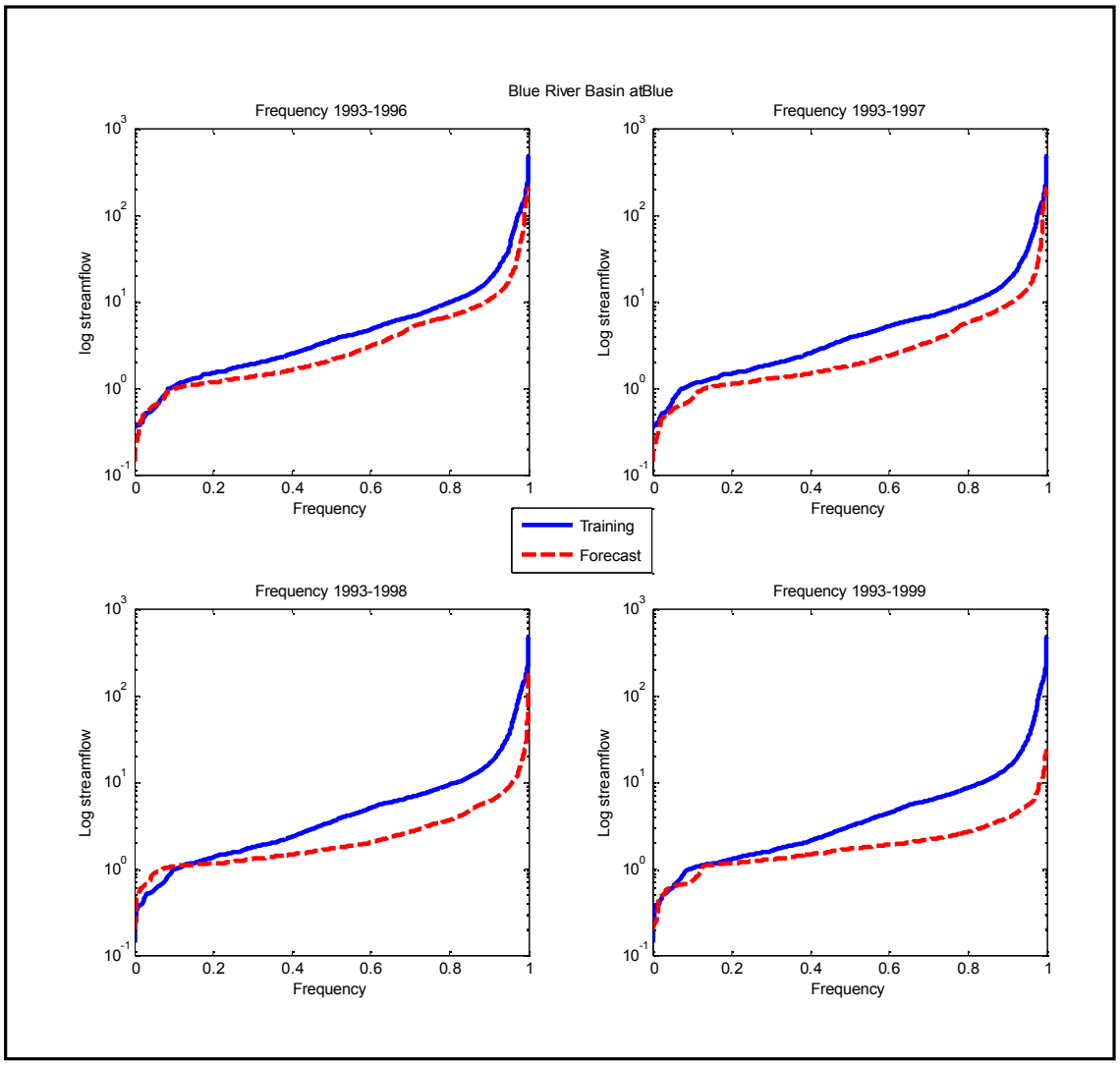


Figure 5: Comparison of streamflow frequency during training period and forecast period

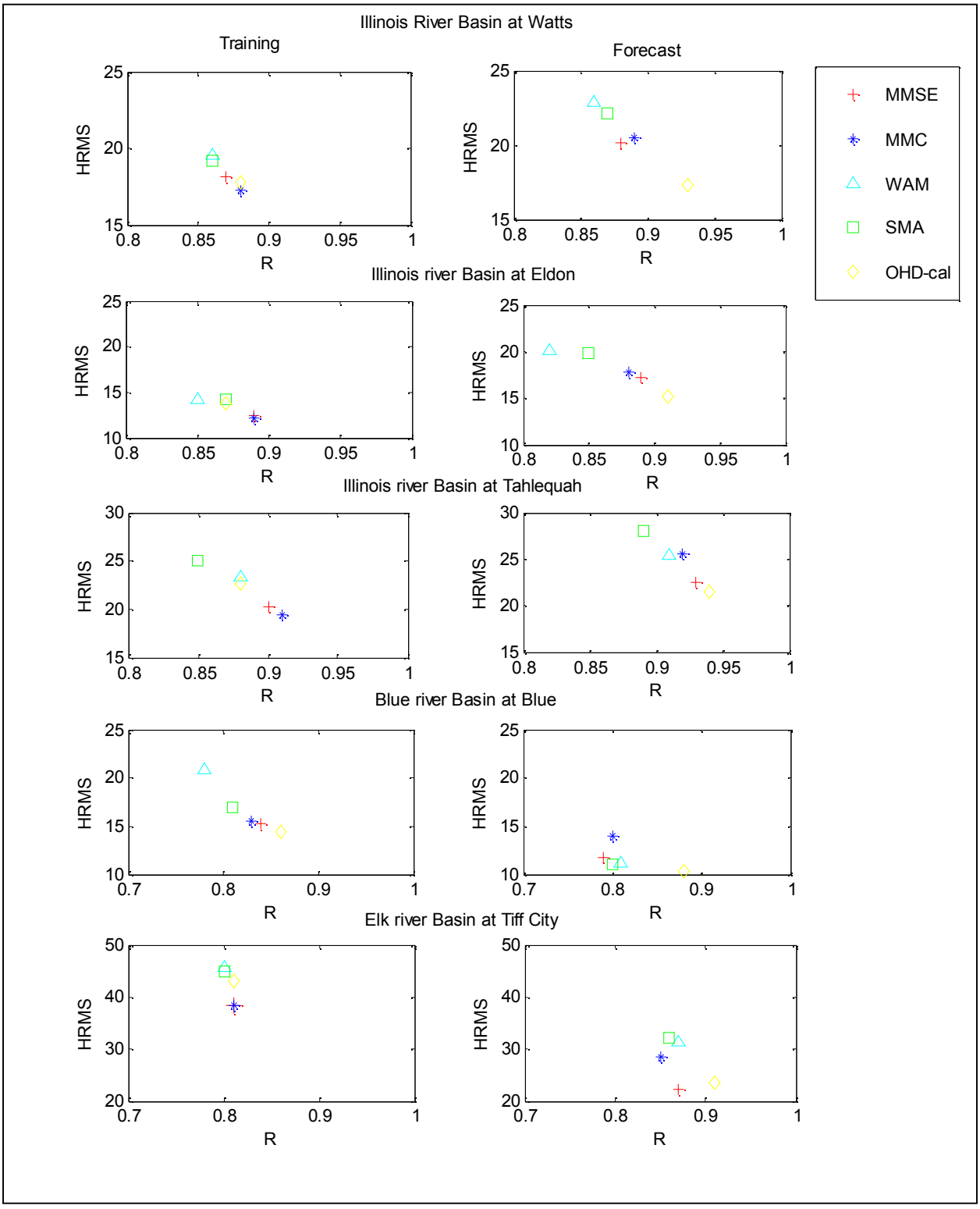
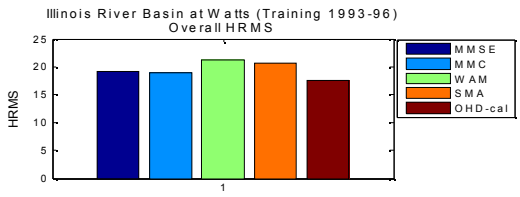
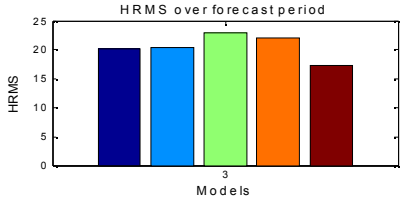
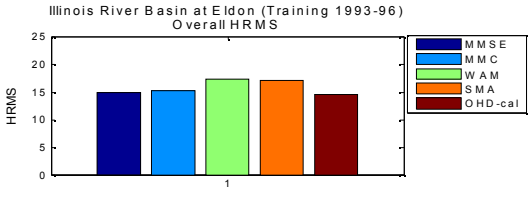
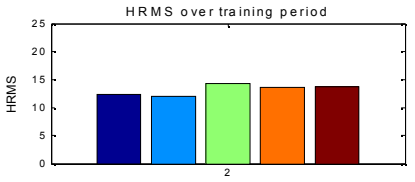


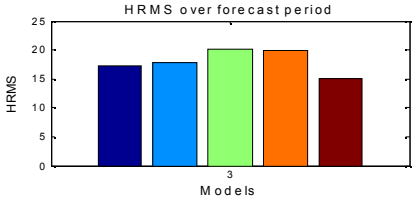
Figure 6. Hourly root mean square error versus Pearson Coefficient for all model combinations (MMS, MMC, WAM and SMA) as well as the best performing calibrated model (OHD-cal) for all the basins (the closer to the bottom-right corner the better the model)

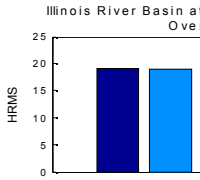


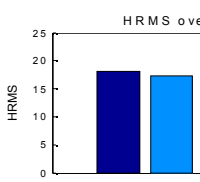


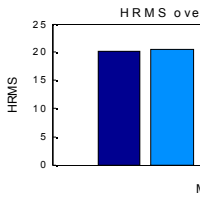


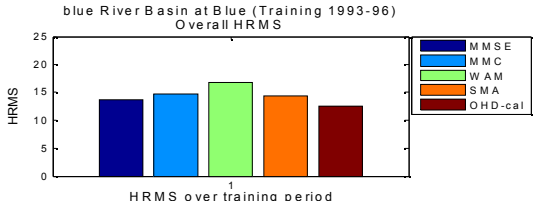


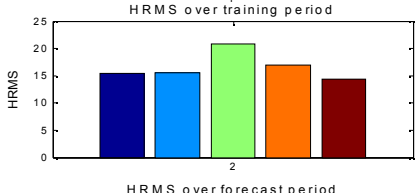


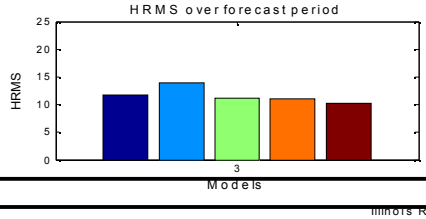


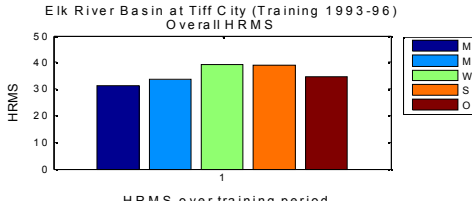


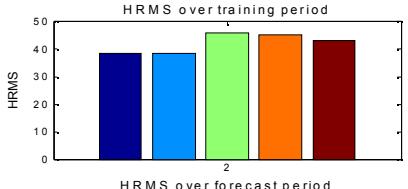












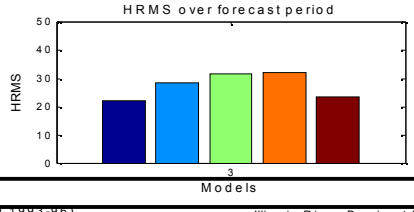
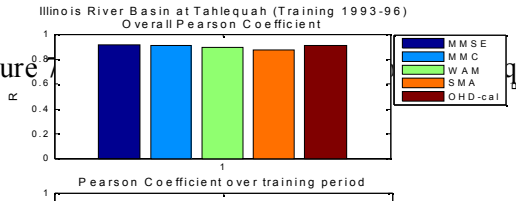
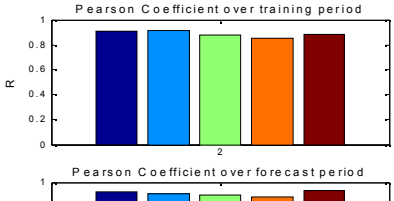
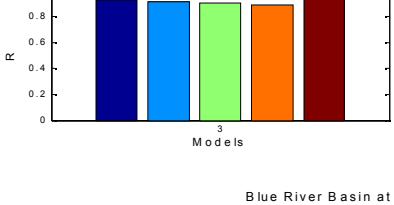
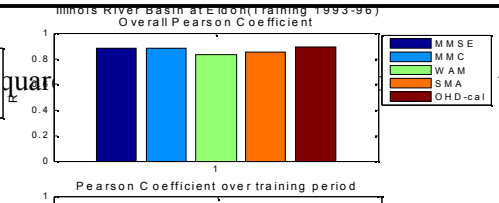


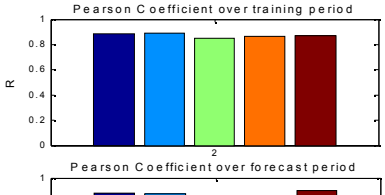
Figure 7 [illegible] quar[illegible] the mu[illegible]

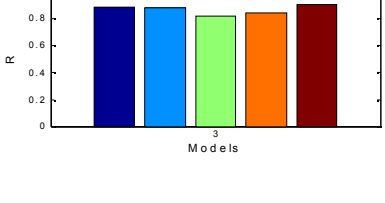


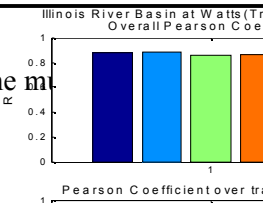


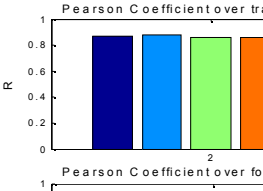


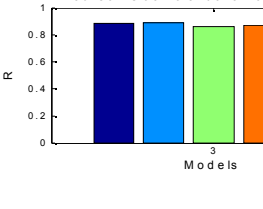


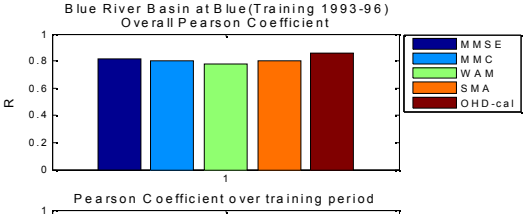


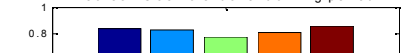


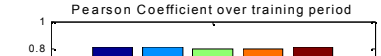












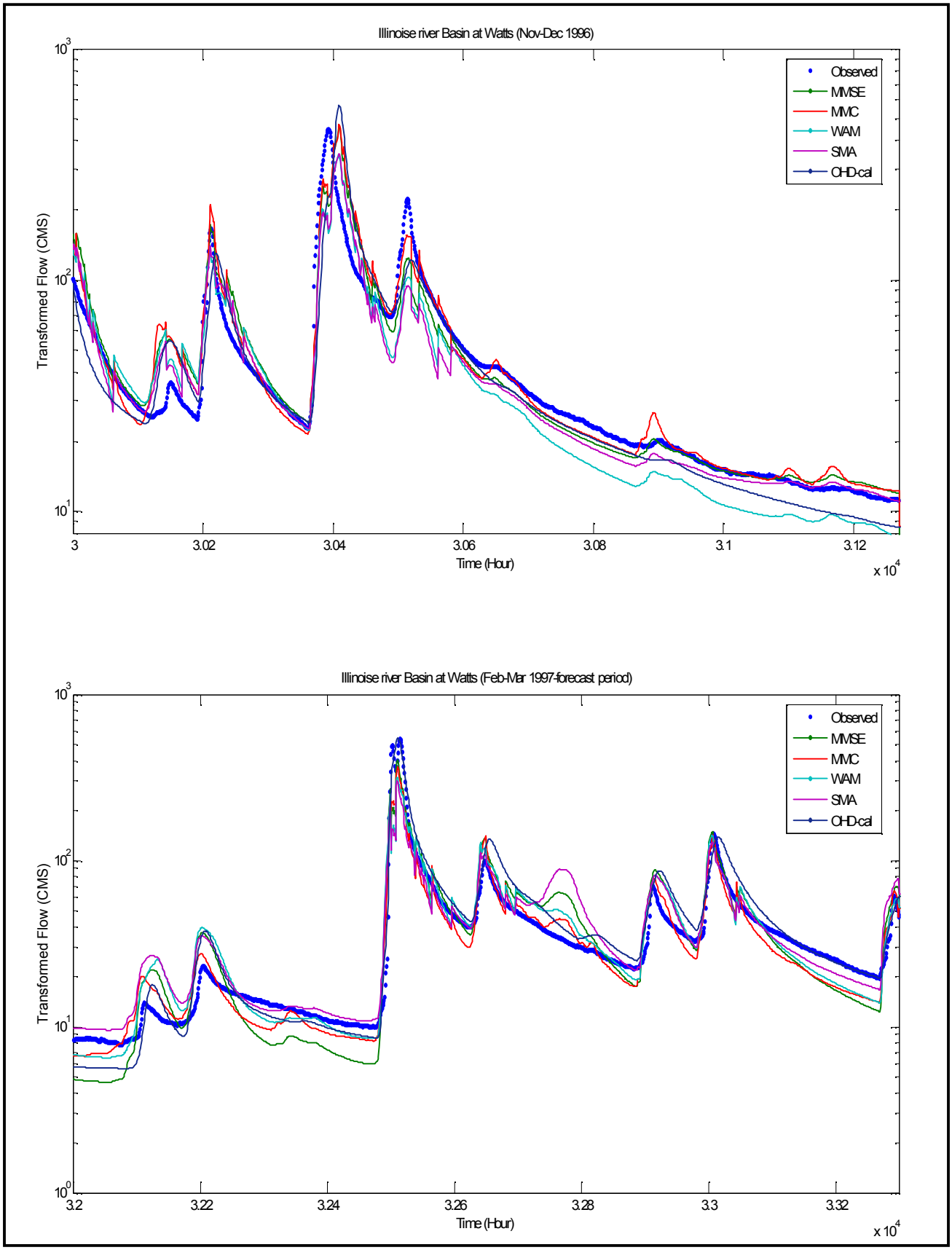


Figure 9. Excerpts of flow simulation results for Illinois River basin at Watts during training and forecast period, illustrating the performance of all combination techniques as well as best calibrated model compared to observed flow

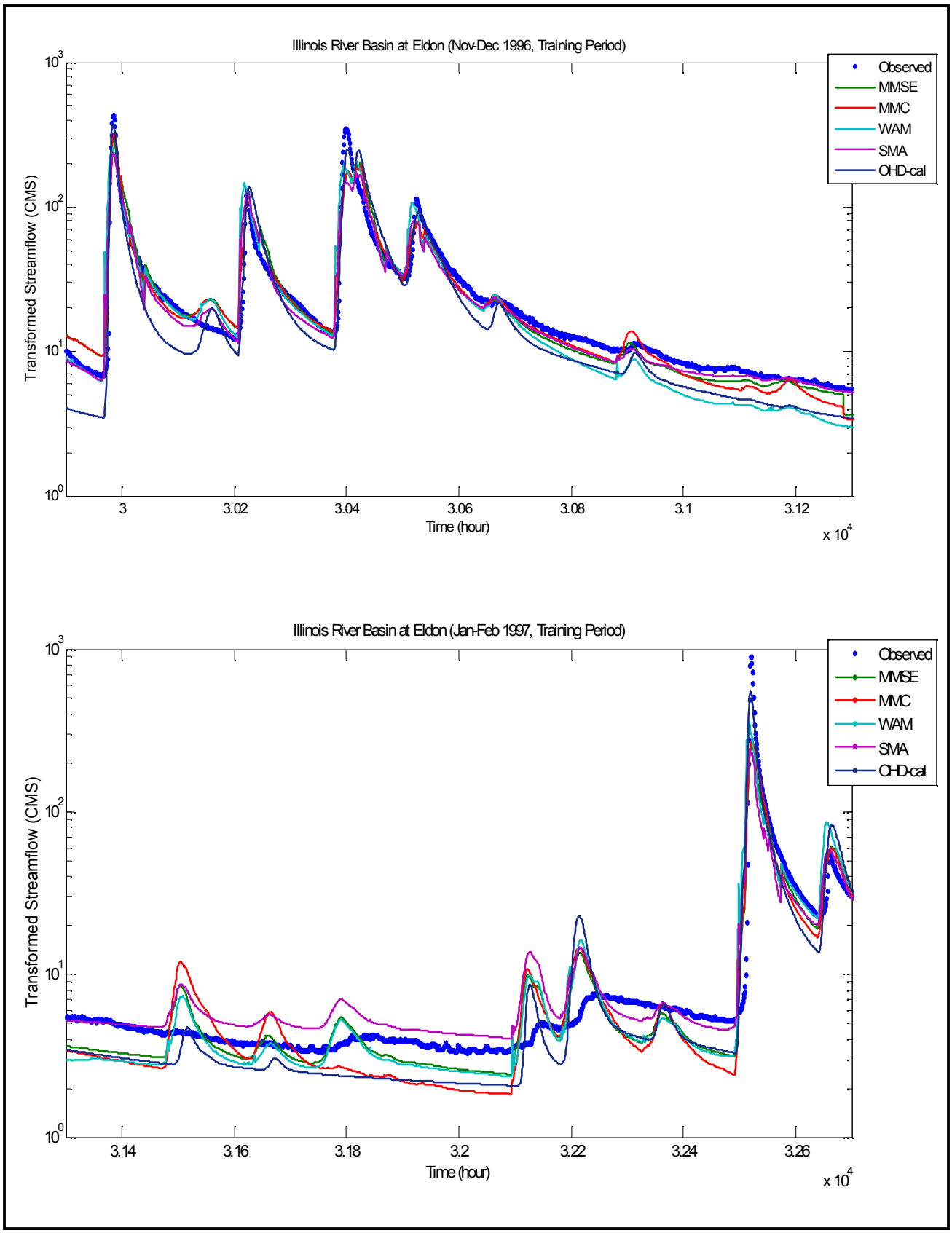


Figure 10. Excerpts of flow simulation results for Illinois River basin at Eldon during training and forecast period, illustrating the performance of all combination techniques as well as best calibrated model compared to observed flow

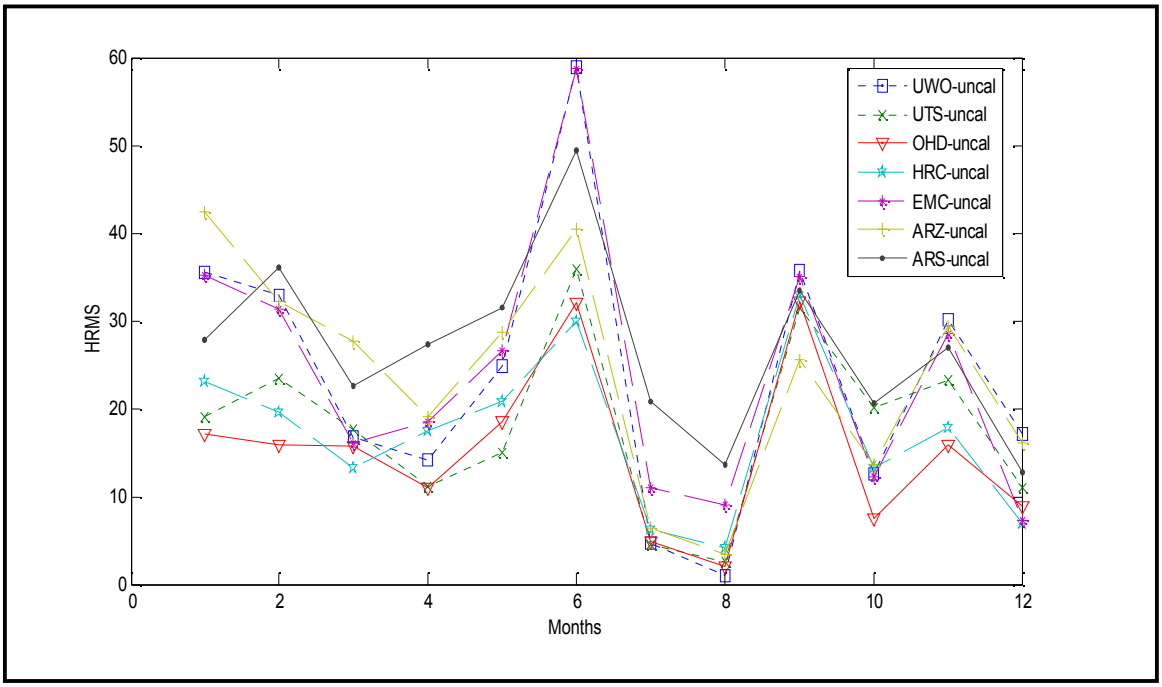


Figure 11. Monthly HRMS of uncalibrated member models for Iliinoise River Basin at Eldon

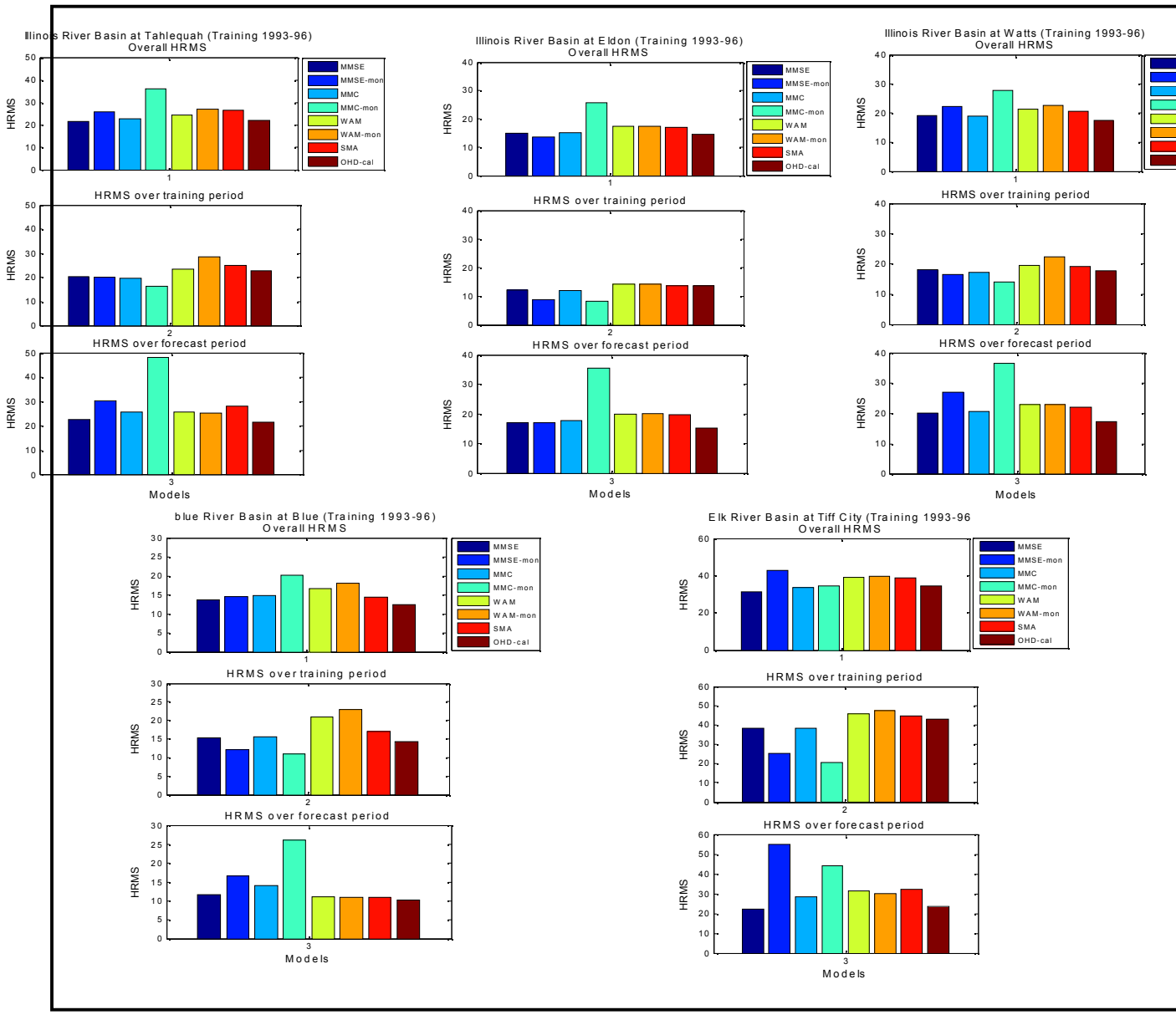


Figure 12: Hourly Root Mean Square error of different combination methods including monthly combination techniques fo

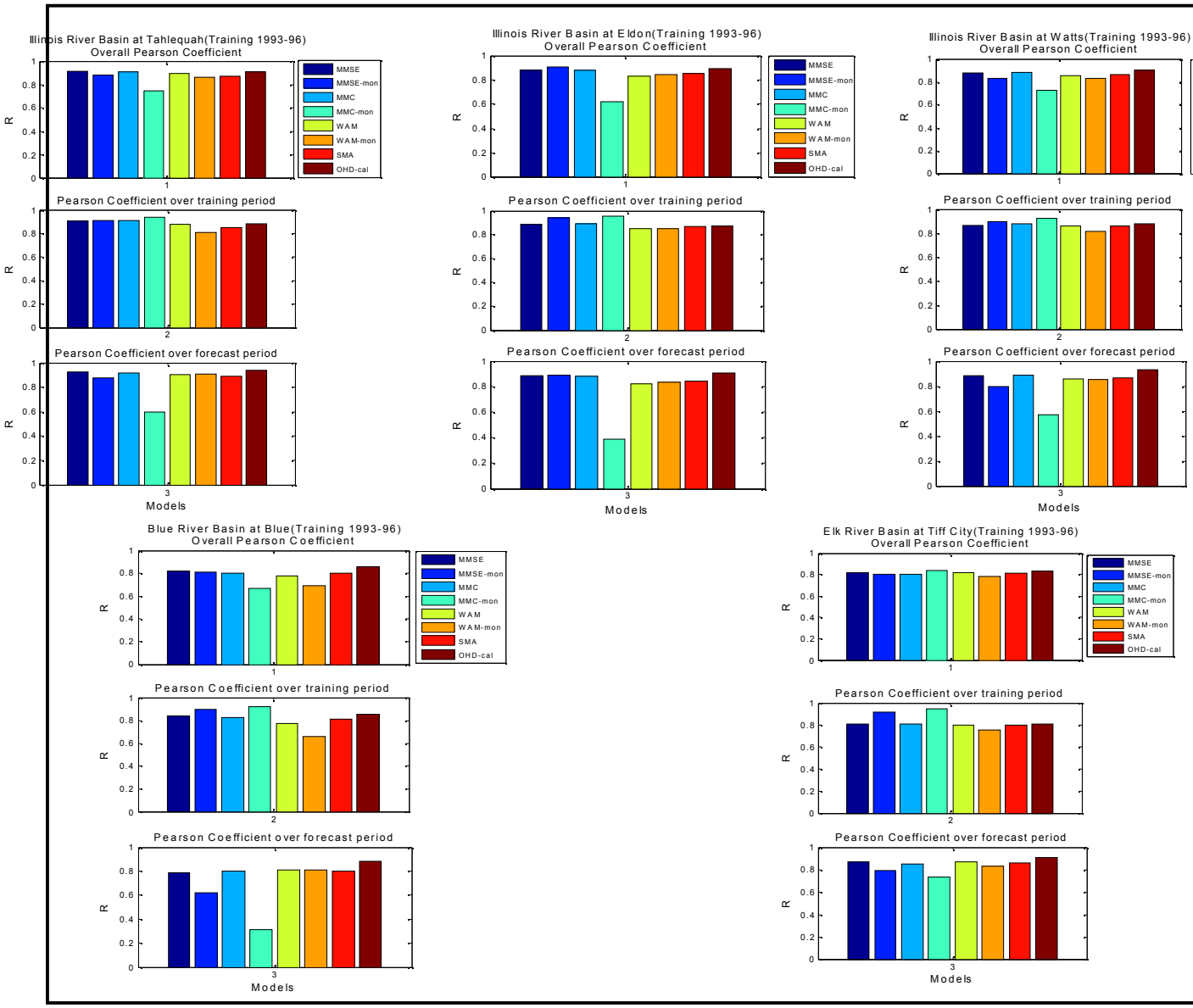


Figure 13: Hourly Root Mean Square error of different combination methods including monthly combination techniques

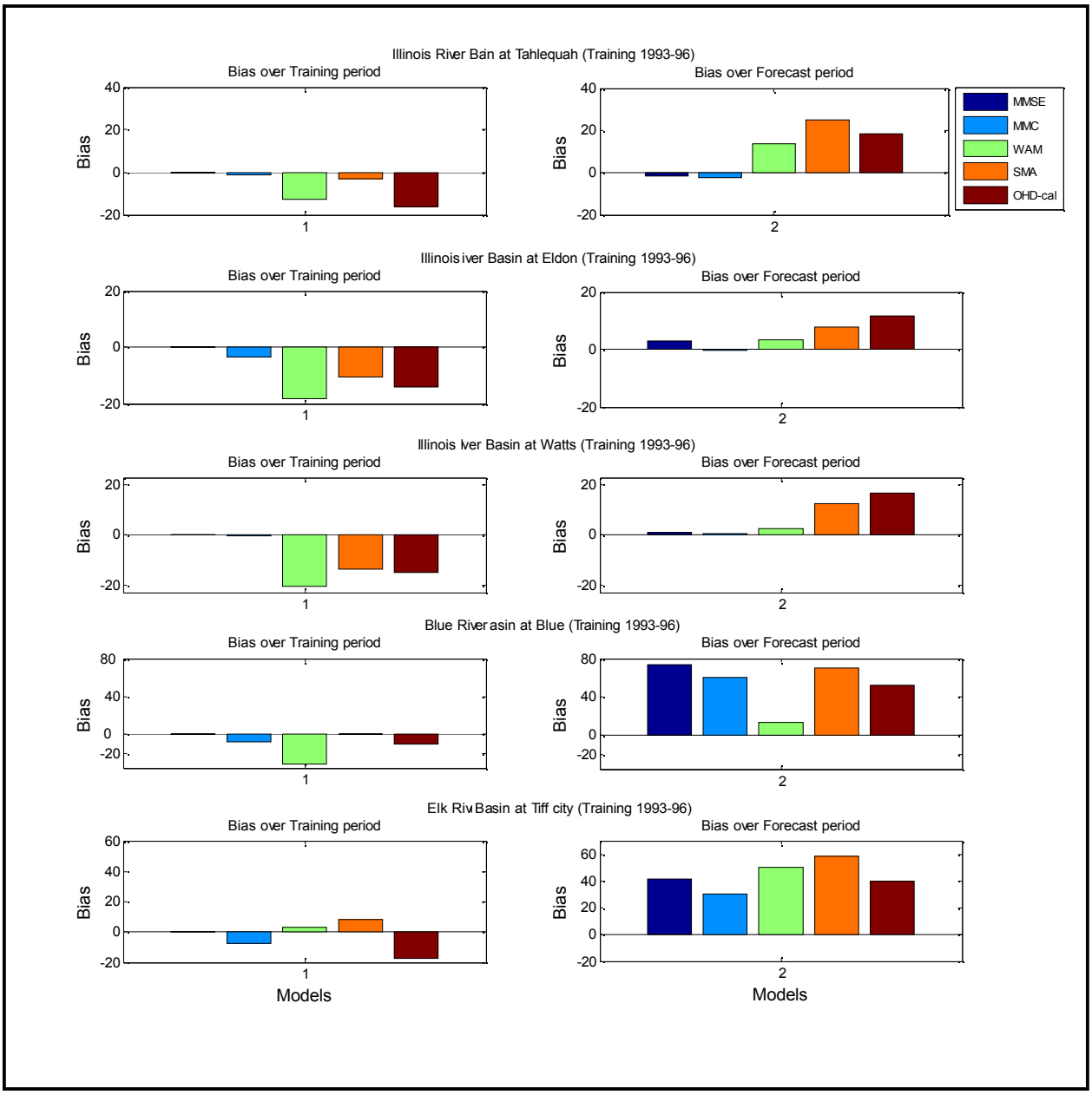


Figure 14. % Bias during Training anf forecast period for all the basins

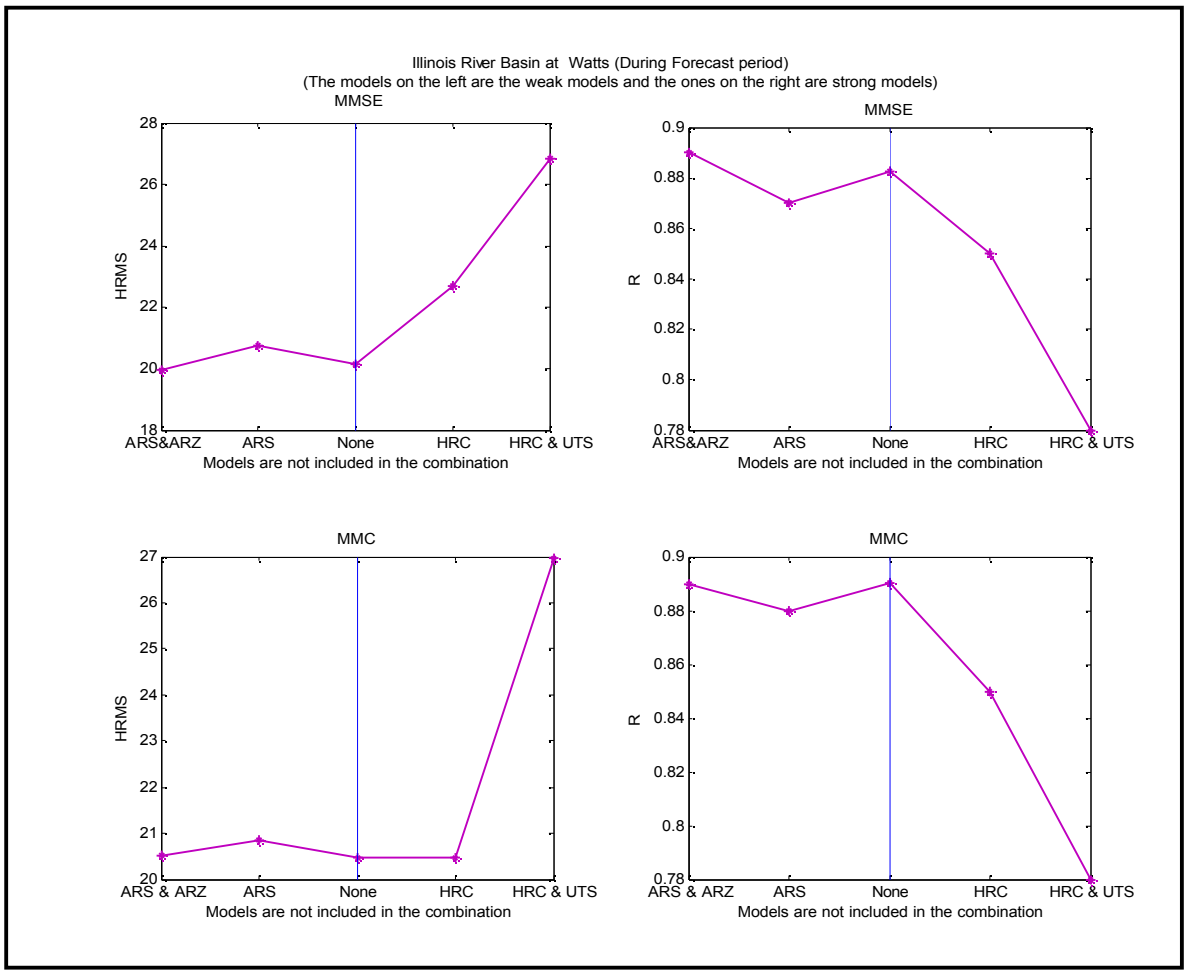


Figure 15. Number of models needed in the multi- model set for the best performance of combination

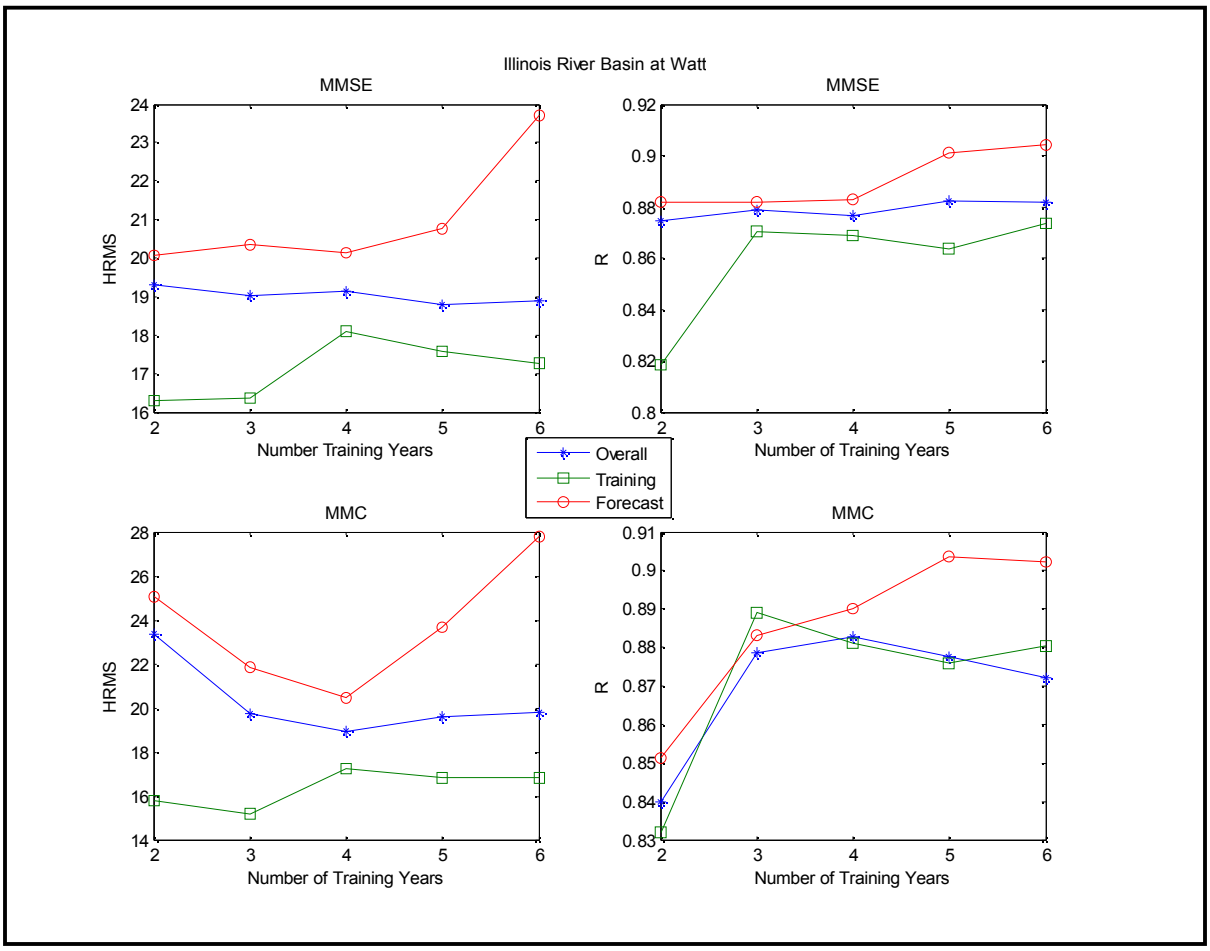


Figure 16. The Least length of training period for optimal performance of MMSE